`DEPARTMENT OF HEALTH & HUMAN SERVICES

PUBLIC HEALTH SERVICE
Food and Drug Administration
Center for Biologics Evaluation and Research
1401 Rockville Pike
Rockville, MD 20852

Memorandum

Division of Clinical Trial Design and Analysis
HFM-576

Date: December 12, 2002

Subject: Advisory Committee Briefing
BLA STN 103979 / 0

Topic: Genzyme Historical Dataset for Use as Study Control Group

Genzyme has submitted a marketing application requesting Accelerated Approval for their galactosidase product for treatment of Fabry Disease. In support of this development program, Genzyme initiated a placebo-controlled, randomized study of patients with Fabry Disease as the verification study required by the accelerated approval regulations. This study is ongoing. Genzyme now proposes to convert this placebo-controlled study to an open label, single arm study by crossing all placebo patients to active treatment and continuing the study. Comparison of results on progression of renal disease would be with an historical database as the control for comparison.

In order to support this modification, Genzyme performed a data-collection study to better define the natural history of patients with Fabry Disease. Genzyme has proposed that the resulting database provides an adequate basis to provide a historical control for use as a comparison group to a treated group of subjects in an open label, single arm study. Genzyme now proposes that this single arm study would be their Phase 4 verification study after receiving an Accelerated Approval.

This document will review the data submitted by Genzyme for this purpose. An introductory section provides a perspective on verification studies in the setting of Accelerated Approval. Other potential uses of this dataset, for example, as a guide to better estimating the sample size required in a randomized, placebo controlled study will not be addressed in this document.

This document concludes with a brief overview of a new, but not fully described, proposal from Genzyme for a different method of employing these data as a comparison group.

Table of Contents

Verification Studies in Accelerated Approval of Products

Accelerated approval is a form of approval to market a product (a drug or biologic) that has been studied for its safety and effectiveness in treating a serious or life-threatening illness and that provides meaningful therapeutic benefit to patients over existing treatments. Among the regulation stated specific circumstances when accelerated approval may be employed is that FDA may grant marketing approval for a biological product on the basis of adequate and well controlled clinical studies establishing that the biological has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit. Approval is subject to the requirement that the applicant study the biological product further to verify and describe its clinical benefit (hereafter referred to as a verification study). Such studies would usually be already underway at the time of the approval. These studies must also be adequate and well controlled, and the applicant shall carry out such studies with due diligence.

Under accelerated approval regulations, the FDA may withdraw approval if the verification study fails to verify clinical benefit or if the applicant fails to perform the required study with due diligence.

Several aspects of the requirements of the postmarketing study bear emphasis. First, the verification study's goal is to determine whether the drug truly provides clinical benefit. Directly establishing the validity of the surrogate as a predictor of clinical benefit is not required. The applicant is required only to study the product's clinical efficacy in a manner that provides substantial evidence of the product's clinical benefit. While an applicant may collect further information on the surrogate along with the clinical efficacy data, and analyze this data to assess, and potentially validate the surrogate, this is not a requirement of the regulations.

Second, to achieve the goal of determining the clinical benefit an providing substantial evidence, the verification study should have several characteristics. The study should be adequate and well-controlled. The study should be feasible to conduct and avoid type I error. Importantly, the design should have adequate power to verify the drug's clinical benefit. If an inconclusive study result occurs, patients and physicians will be left not knowing whether or not the drug does provide benefit, and FDA will need to consider the withdrawal from marketing of a drug that might, in fact, be beneficial.

An additional consideration is that orphan drug designation does not alter the legal requirement for substantial evidence of effectiveness. Although Genzyme's product has been granted orphan drug designation, the same requirement for substantial evidence of effectiveness applies as for other products.

Design and Conduct of Historical Data Collection and Disposition

Data Collection Design

Genzyme prospectively designed a protocol for the collection of the historical data submitted to the FDA. This study was titled "Epidemiologic Study of the Natural History of Fabry Disease",

Protocol # AGAL-014-01.   This protocol was finalized on March 22, 2001, and was conducted for Genzyme by a contract to Abt Associates Clinical Trials (AACT), a contract research organization (CRO).

Briefly, potential study sites were contacted and participation was discussed. At sites willing and able to participate, patient records were examined. All patients who had been given a diagnosis of Fabry disease which was not later rescinded were eligible. Male and female patients, of any age, and of any time frame of diagnosis (i.e., not limited to patients first diagnosed within any specific time period) were eligible. Patients were excluded who had a concurrent disease that would confound interpretation of renal function information, or if they had some other serious disease at the time of diagnosis. Patient records were abstracted from the time prior to the period of receiving any enzyme replacement therapy if that patient had participated in a study of enzyme replacement from any manufacturer. In addition, patients or another authorized party were required to give consent for the collection of the information from the medical record. Screening logs of all patients considered for the study at each site would be maintained.

The data collected focused upon measures of renal function and renal related adverse events, cardiac function and cardiac adverse events, stroke and mortality. Demographics, date and basis of Fabry diagnosis, and other information characterizing the disease were abstracted as best as possible from the information available.

The CRO performed all abstraction of data at the clinical sites. Analyses of the data were performed largely under the direct supervision of and/or by Genzyme personnel.

Because of the need to contact and gain consent from each patient, and the need to carefully review all records in the patient's medical history at each site, the study was not completed as rapidly as initially expected. The final study report on all aspects of the data collected has not yet been submitted to FDA as of the date of writing this review. However, Genzyme has conducted two interim analyses of the data. This review is based on the data contained in the second interim review. This interim has data that are thought to be largely complete and final with regards to renal function. Other portions of the data (e.g., cardiac, cardiovascular) may not be as complete, and have not been analyzed or submitted to FDA as of the date of this review.   The second interim report is based on the study data with a Genzyme cut-off of June 5, 2002.

Comment: Genzyme has subsequently completed their full analysis and report of these data and the final report has been submitted to FDA. Changes to the dataset were, as expected, minimal. The "Qualified Patient" dataset changed from 103 patients to 104 patients. Few other changes were noted by Genzyme. Genzyme did not recalculate or submit many of the specialized, FDA requested, analyses described in this memorandum. However, such small changes in the dataset are unlikely to change the results in any significant manner. Consequently, this memorandum remains largely based on the "second interim" review of the dataset, but is fully applicable to the final dataset as well.

Site, Patient and Data Disposition

There were 51 sites invited to participate, located throughout North America and Europe.   Of these, 27 sites chose to and were able to participate, 19 in the US, 5 in Canada, and 3 in Europe. No explanation of the reasons for not participating was provided to FDA.

The second interim analysis contains data through June 5, 2002. As of that date, all patients had been enrolled and had all data abstracted. However, most but not all data-verification queries had been completed. For purposes of facilitating the review and discussion of the primary interests of Genzyme, priority had been given to the renal function and renal adverse event data for verification. Therefore, Genzyme states they believe that there will be only minor, if any, changes to the database in these areas prior to the finalization and completion lock of the data.

All identified Fabry patients were attempted to be contacted at each site. Genzyme has not completed collection and analysis of the screening logs as of yet, but has provided an interim summary of the screening results. This interim review contained 742 identified patients distributed among the 27 sites. Of these, 431 (58%) gave consent to have their records examined and abstracted. A substantial number, 264 (36%) did not respond to inquiries. Only 3% replied they were not interested in participating and 1% were identified as having died by the time of the historical review. There were also 2% who were known to be not contacted due to known incorrect address, without correction available.

There were 447 total patients in the database, 279 (62%) male and 168 (38%) female.

Genzyme Analysis

Definition and Creation of a Data Subset

Genzyme has the primary interest in using these data as an historical control group for an open label, single arm study of enzyme treated subjects. To support this use, Genzyme divided the historical database patients into two categories, those that have at least a portion of their medical history which includes a timepoint when the patient would have been eligible for inclusion in Study 008 (the current randomized, controlled study), and those that have no point in time of the medical records when they would have been eligible. Genzyme has named the patients whose data indicated eligiblity at some point in their historical record the "Study 008 Qualified" patient group (or: Qualifier group).

Qualifiers would need to have a point in the historical record when documentation indicated they would have met all inclusion criteria, and not have had any exclusion criteria.   At the first documented time of that occurring, the patient was designated as having qualified, and that calendar date is regarded as the date of qualification.

The major eligibility criteria for Genzyme Study 008 are:
        Age >= 16 yo, male or female
        Diagnosis of Fabry Disease, and clinical presentation consistent with the diagnosis

Documented   plasma αGalactosidase activity of < 1.5 nmol/hr/mL or
  Leukocyte αGalactosidase activity of < 4 nmol/hr/mg
Renal impairment, as documented by:
  Serum creatinine of 1.2 – 3 mg/dL   or
  Estimated creatinine clearance < 80 mL/min (if serum creatinine < 1.2 mg/dL)
Disqualification if:
  Currently on dialysis or scheduled for renal transplant, or in acute renal failure
  Cerebrovascular events within prior 3 months
  MI within prior 3 months, current unstable angina, CHF of Class III or IV
  Other serious medical conditions, including diabetes mellitus, cancer, etc.

Of the 447 patients in the database, documentation supported only 115 meeting the inclusion critiria, and 12 of these had exclusion criteria adverse events, so that only 103 patients (23%) remained in the Qualifier population.

Comment: It is important to bear in mind that the patients records in the historical data collection were not originally recorded for purposes of identifying a "Study 008 Qualified" period of time. A reasonable likelihood is that most Fabry patients experience a period of time when they would qualify for Study 008. However, the limited numbers who are classified as having such a period of time in the historical database highlights the non-prospectively planned nature of the medical recording of information on these patients at the participating sites. Many of the patients had primary care physicians outside of the participating medical centers, so much of their medical care is not recorded in the materials that were available for the abstraction procedure.

The reasons for not qualifying are illustrated in the following table:

Qualifier Criteria Disposition among 447 total patients					
Criterion	Qualifying		Not Qualifying		Missing
Age	>= 16 yo		< 16 yo		
	385		24		38
Galactosidase activity	Leukocyte < 4	Plasma < 1.5	Leukocyte >= 4	Plasma >= 1.5	
	126	119	76	62	64
Serum Creatinine	1.2 - 3	<1.2 but CrCl < 80	All < 1.2 and Cr Cl > 80	All >= 3	(no Cr values)
	139	72	186	12	38
Age & Galactosidase	212		235		
All 3 criteria	115		332		
All 3 criteria & no excluding event	103		344		

"missing" age permitted only in subjects without any recorded creatinine values, otherwise age was sought; CrCl = creatinine clearance

For the 23% of patients who had some data permitting qualification by Study 008 criteria, the start of the record was deemed the date of qualification, and all subsequent creatinine values were deemed as belonging to the "qualified" period until a disqualfying adverse event occurred (e.g.,

dialysis, renal transplant, stroke, MI, diagnosis of cancer, etc.) or until the patient received enzyme replacement treatments within a clinical trial.

All further analyses in the Interim Analysis have focused on this Qualifier subset of 103 patients. These analyses include all creatinine values recorded on or after the date of qualification up to the date of disqualification, or the last recorded value if no disqualifying event occurred.

Demographics and "Baseline" Characteristics of Qualified Subset

Genzyme presented the following information regarding the 103 patient qualified subset:

Characteristics of the Qualifying Subset at the Date of Qualification ("Baseline")		
Age	mean	37.3
Serum Creatinine	mean	1.5
	median	1.3
Weight, kg	mean	70.2
Height, cm	mean	174.2
Gender	male	97 (94%)
Ethnicity	Caucasian	87 (84%)
	Black	3 (3%)
	Hispanic	8 (8%)
	Other	4 (4%)
Blood Type	B antigen	11 (11%)
	No B antigen	56 (54%)
	Unknown	36 (35%)
Plasma αGal Activity nmol/hr/mL	n=55, mean	0.7
Leukocyte αGal activity nmol/hr/mg	n=48, mean	1.29

Extent of the Historical Data

For the 103 total patients in the qualified subset, there are 583 creatinine values. These datavalues are not distributed uniformly among the patients. Many have relatively few values. There are 18 with only one creatinine value within the qualified period, and 22 have only 2 creatinine values, leaving only 63 patients with 3 or more creatinine values (see following table).

Distribution of number and duration of creatinine values among Qualified patients					
# values	#patients	Initial Qualifying Creatinine ("baseline")		Duration of Creatinine Observations (yrs)	
		mean	median	mean	median
1	18	1.57	1.3	0	0
2	22	1.46	1.29	1.45	0.44
3	9	1.58	1.4	2.24	1.44
4	9	1.37	1.2	2.26	1.1
5	9	1.45	1.4	1.84	0.82
6	8	1.42	1.3	3.66	2.32
7	2	1.15	1.15	3.57	3.57
8	7	1.39	1.4	4.16	2.94
9	3	1.33	1.3	7.17	7.54
10 - 29	16	1.54	1.35	7.99	6.85

Proposed Use of the Historical Data

Genzyme proposes that endpoint for the the prospective agalsidase treatment study be the proportion of subjects reaching a specific renal progression criterion within a fixed time period of starting enzyme replacement treatments. This observed result would be compared to the value in the Qualified Subset of the historical data. Genzyme has discussed in their submissions criteria of both a 33% rise within 2 years of beginning agalsidase treatment, and of 50% rise within 3 years. While the 33% rise criterion is the criterion used in the current randomized form of the study, the 50% rise criterion is a more clear determination of an important decline in renal function. At the time of this writing, Genzyme's focus is on a proposal of incidence of renal progression defined as a 50% rise in creatinine within 3 years. The 3 year period is an arbitrary selection so that the trial is not unreasonably long, yet allows a reasonable period of time to observe significant decline in renal function. The ensuing discussions in this document will focus upon this criterion. It should be noted that selection of this criterion and timeframe for renal progression is not derived from the historical dataset.

Genzyme proposed to use the historical data to develop an estimate for the number of patients expected to reach the progression criterion within 3 years of "qualification", and use this value as the control value for comparison to the clinical events observed in treated patients. The historical data and analyses should be judged on the ability to provide a reliable and valid control comparison for a prospective single-arm study.

Genzyme's Proposed Method

In order to develop the estimate of renal progression from the historical data, Genzyme posits that renal dysfunction in Fabry patients occurs in a manner such that the logarithm of creatinine values are linear over time. For the proposed criterion of a 50% increase within 3 years, slopes that will meet or exceed this increase will have a minimum slope of :

$$\{\ln(1.5\mathrm{creat}_{bl}) - \ln(\mathrm{creat}_{bl})\}/3 = 0.135 \ [\ln(\mathrm{creat})/\mathrm{yr}].$$

Given the assumption of linearity, the estimated natural history progression rate was calculated as follows.

The historical creatinine data (in logarithm transform) was used to calculate an estimated slope (units of ln(Creat.)/year) for each patient by application of a linear random effects model with empirical Bayes estimation.

Second, the 103 estimated slopes are rank ordered, and the fraction of all slopes that are greater than 0.135 is the proportion of patients estimated to have renal progression within 3 years. The distribution of slopes that Genzyme calculates among the Qualified patient dataset is shown in the following histogram.
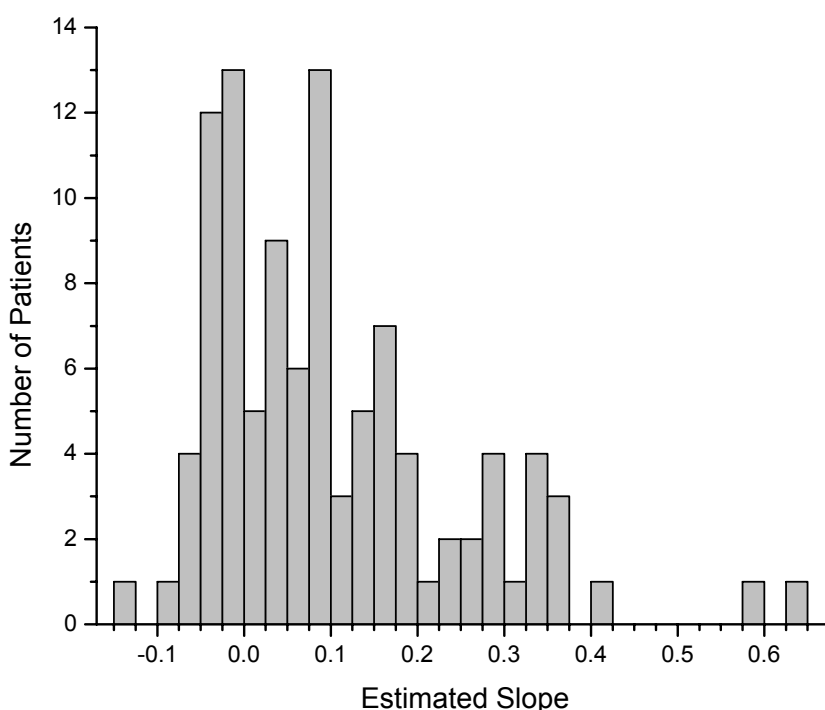
An alternative display of this same analysis result is in the form of a cumulative distribution of the percentage of the analyzed population (i.e., a curve of the percentage of patients with a slope less than or equal the corresponding slope value). Such cumulative displays better allow for comparison of different sensitivity analyses, and more readily illustrate changes in the percentage with slopes less than (and consequently, more than) the critical value of 0.135.
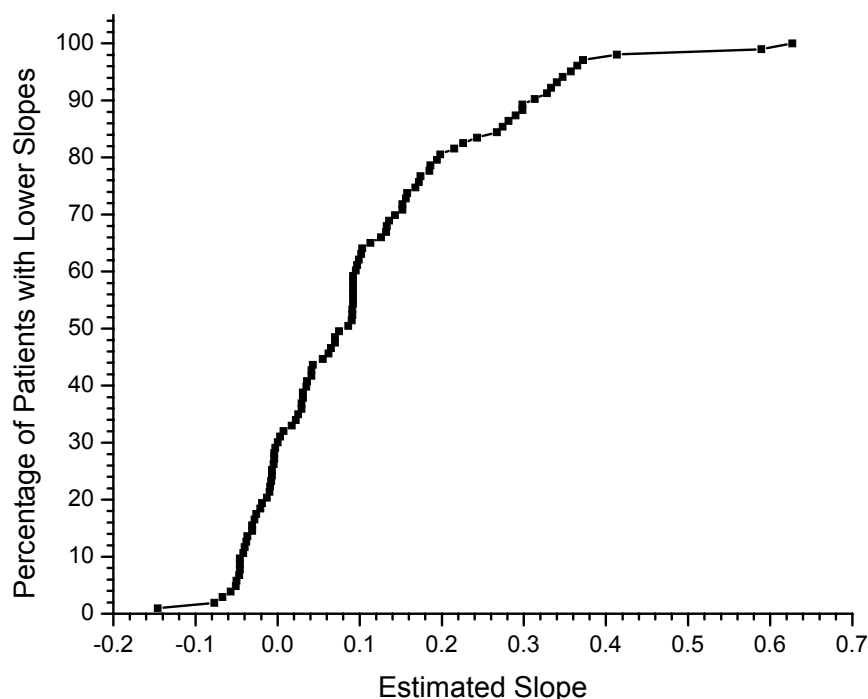
This figure illustrates that 68% (70 of 103) of the historical database patients have a slope of less than 0.135, so that the predicted renal progression rate is 32% (33 / 103).

A notable feature of this procedure is that because of the reliance on the empirical Bayes method, slope estimates can be made even for patients with only a single creatinine value in the dataset (18 of the 103). To explore if this aspect markedly shifts the estimate of proportion with projected renal progression, Genzyme conducted the modeling procedure in reduced datasets of only qualified patients with at least 2 creatinine values (n=85 patients), and in those with at least 3 values (n=63 patients). Genzyme reports that the point estimates for proportion with progression were not changed.

Comment: This not an entirely unexpected observation. The empirical Bayes aspect of the modeling acts to reduce the variability of the estimated slopes by influencing each fitted slope to be similar to the entire body of data. The patients in the body of data are not evenly weighted, but weighted in relation to the number of datapoints available for each patient. Thus, patients with many datapoints are considered more informative than those with few, and those with one data point contribute little to the model's construction of the expected slope distribution. These patients are the most influenced by the modeling procedure to conform to the remainder of the patient data. The patients with a single creatinine value are assigned a slope value in a manner that is entirely linearly related to the logarithm of the initial (and only) creatinine value, conforming to the relationship seen in the dataset as a whole (see section on Linearity, below). Thus dropping out the patients with only one or even two points only is not expected to change the distribution significantly; the slope estimates for these patients were the most heavily influenced to conform to the patient population as a whole. Note, however, that some estimates of confidence intervals will be affected by inclusion of the single-value, or two-value patients. Single or two-value patients will contribute slope values and outcome prediction that are given equal weight to slopes or outcome

from patients with many values when the sample size (number of patients) is applied to the estimation of uncertainty for the confidence interval.

Genzyme has provided "confidence intervals" for this estimate of renal progression rate. Genzyme's method to perform this calculation has been to calculate a confidence interval (C.I.) around the mean slope value (using variance from the variance-covariance matrix of the random effects with empirical Bayes estimation model). Genzyme then applies this same C.I. width to the slope value 0.135 to define an interval (0.135-W, 0.135+W). Using the cumulative slope-distribution curve this interval is then translated to the fraction of patients with slopes greater than each interval edge. Genzyme states, for example, that the 90% confidence interval around the proportion of interest is 24% to 39%.

Comment: However, Genzyme's may have employed an invalid method for calculation of confidence intervals (or "tolerance intervals"). Genzyme calculates the standard error (uncertainty) of the mean slope from the patient data. Genzyme then applies this standard error to prespecified slope of interest (0.135). This method assumes that the standard error around the slope value 0.135 is the same as the standard error around the mean slope. The method is focused on calculating the uncertainty in a slope, and then using the empirical curve to transform that slope uncertainty to an interval in percentages. However, since there is no uncertainty regarding the slope of interest (it is 0.135 by prospective selection) the meaning of a standard error interval around this value is unclear in this context.

A proper method for CI calculation will be a method that directly addresses uncertainty in the percentile at which slope = 0.135 occurs, rather than the uncertainty in the mean slope. Additionally, any method that attempts to perform this calculation will need to consider the dataset basis for the slope distribution curve. Many methods will weight each data value equally. However, patients that have only 1 creatinine value may need to be considered to have more uncertainty in their slope estimate than those with multiple values. Unlike the observation above regarding the point estimate based on dropping patients with few datavalues, confidence intervals of the proportion of patients with slope > 0.135 are likely to be different for the datasets of patients with at least 2 values per patient (n=85) and those with at least 3 values per patient (n=63).

For example, if the contributions of measurement errors, modeling errors, and other errors in the assignment of slopes are ignored so that all estimated slopes are assumed completely accurate, and the population were assumed completely representative of historical subjects, a simple approach is to calculated a confidence interval on the incidence of 33 of 103 (32%) with slope > 0.135. The 95% C.I. for this incidence is 23% to 41%. Restricting the modeling to patients with 3 or more datapoints also gives a 32% point estimate, with a CI of 20% to 43%. However, errors in measurements will add to uncertainty, and sampling biases that may make the collected population unrepresentative will bias this estimate.

Genzyme Assessment of Stability of the Projected Rate

Empirical Estimate

Genzyme examined the Qualified patient database to tabulate the empirically observed rate of renal progression in those patients who have a data value available at approximately 3 years. The

empirical progression rate was 7/17 patients or 8/18 patients (41% and 44%, respectively) depending on use of a window for the critical creatinine value of either 2.5-3.5 years, or 3-4 years. Although markedly different in the simple point estimate, these rates are not substantially different from the modeled estimates when taking into consideration the very limited numbers of patients this analysis can include. However, the very limited numbers of patients that can be included in this estimate limit the strength of this comparison.

Subdivided Dataset Estimates

Comment: The final result of progression rate from the data modeling method is dependent upon the patients included in the analysis, and Genzyme's stepwise interim analyses illustrate this. The complete dataset evaluated in this document is actually the second analysis performed by Genzyme. Genzyme had conducted a previous interim analysis with inclusion cutoff of patients available on an earlier cutoff date. At this first interim analysis only 43 patients were in the qualified dataset. The estimated event rate from this dataset was 40%, as there were 17/43 patients with an estimated slope >= 0.135. This is substantially different than the estimate of 32% from the current dataset, which includes the original 43 patients. Notably, among the 60 qualified patients added for the second analysis, only 16 were estimated to have a slope > 0.135, for an estimated projected progression rate of 27% in the second portion of patients. Thus, the event rate estimate is dependent upon the specific patients included in the modeling procedure. The projected progression rate is not stable even to the two portions of this dataset.

Covariate Effects

According to the sponsor, age, gender, weight, blood type, and plasma galactosidase do not substantially alter the estimated event rate (i.e., the fraction of slopes >= 0.135) when included in the model as an offset value for each patient. Genzyme has not thoroughly evaluated the relationship between these covariates (or others) and the estimated slope for each patient.

Comment: CBER addresses this topic in a later section.

Selection of ln(creatinine) and linear fits for modeling

In discussing the basis for the use of logarithm of creatinine in the modeling procedure, Genzyme notes that the untransformed creatinine data do not adhere to a linear change over time, but that Genzyme regards the transformed dataset of ln(creatinine) as showing adequate linearity. Genzyme based this conclusion, in part, on a lowess curve analysis ("robust locally weighted regression and smoothing scatterplots"). The stated advantage of ln(creatinine) in that it is convenient, as the slope for any specific relative change in creatinine (e.g., 50% increase within 3 years) is easily calculated (i.e., $\ln(1.5)/3 = 0.135$). Consequently, Genzyme assumed that ln(creatinine) changes linearly with time.

Comment: See section on assessment of linearity later in memorandum for further interpretation of lowess curve.

Genzyme explored other models of the ln(creatinine) data, specifically models with quadratic and cubic terms, in addition to the linear term. The note that the higher order models provide poor predictions of serum creatinine at time periods of 7-8 years and beyond.

Comment: Neither model performance at long periods out nor model performance at time periods preceding the historical data period are relevant to the proposed use of the models. Linear changes in ln(creatinine) over time imply ever decreasing creatinine at progressively earlier times. This is clearly not the case, as all patients had a long period of largely stable creatinine values prior to reaching the period when the value began to increase. Results with quadratic modeling are presented later in this memorandum.

Genzyme did not make goodness of fit comparisons between the different models within the actual dataset. They indicated that when modeling is done with reduced datasets by limiting the duration of observation within each patient (discarding longer term values), estimated overall event rates are generally unchanged and similar between the linear and quadratic models until extremes of data elimination are reached, (i.e., retained data duration so short the models are effectively used to project to timeframes in excess of the retained data). Because the models give similar estimates for the overall progression event rate, Genzyme does not regard goodness of fit as important, and thus the linear model was retained because of ease of use for establishing criteria for the overall progression rate.

Comment: Genzyme's discussion does not adequately examine the issue of whether or not linear models are a good fit to the data. Their submission does provide figures assessing normality of standardized residuals that suggest the quadratic fit has a more normal distribution of standardized residuals than the linear model, but these are not quantitatively compared. Furthermore, since the proposed use for the modeling is not applied on data limited to only less than the time period for the proposed renal progression criterion, the relative sensitivity of the models to such an extreme data limitation is of no practical import. Genzyme does acknowledge that the quadratic terms are significant in the models, and thus are a better fit to the data.

Quality of the linear fit may still be important even if the linear and non-linear models appear to give the same point estimate for the overall progression rate. The distribution of estimated clinical courses may still be different so that the uncertainty in the progression rate is different. Furthermore, the different modeling methods may have different sensitivity to various weaknesses of the dataset, so that the dataset may have greater or lesser robustness to sensitivity analyses with different models.

Issues to Consider in Use of the Historical Data as a Control

This proposed use of a historical dataset as control for prospectively collected data on a group of subjects treated with an intervention raises a number of issues. Some of these are generic to any circumstance of historical controls, and some are specific to the particular method of data use.

The general conditions include:

1)   The assumption that the historical patients are comparable to the patients to be enrolled in the new study.

2)   The assumption that factors external to the patient's intrinsic disease process are comparable to the patients to be enrolled in the new study. Alternative treatments, or other forms of medical management that may have an influence on aspects of the clinical course and may differ over time are a particular concern in this regard.

Items 1 & 2 are needed to ensure that the analysis of the historical patients will allow accurate prediction of the clinical course the subjects in the new study would have had, if they were untreated.

3)   A belief that the analysis of the historical data will allow sufficiently precise prediction of the clinical course of subjects in the new study. This implies that the information in hand regarding the historical patients is sufficiently robust to enable good predictions. Factors that may have an influence on this include how the patients were assessed. If there is considerable measurement error in the assessments, or inadequate frequency of measurement, then the precision of the historical prediction can be adversely affected.

Genzyme proposed a random effects linear model of ln(creatinine) with empirical Bayes estimation of the individual slopes. Necessary particular assumptions include:

1)   That the logarithm of creatinine is sufficiently linear with time in Fabry patients as renal function declines.

2)   That the random effects linear model with empirical Bayes estimation is a sufficiently good method to provide slope estimates.

These issues are, in fact, interrelated. The specific elements incorporated into the modeling process can interact with the assessment of comparability of the two patient groups and the robustness of the dataset.   For example, it may be possible for the patient and disease characteristics which largely influence the disease course to be well understood based on analysis of a highly robust historical dataset. In such a case, the modeling procedure may be able to incorporate this knowledge and be adjusted based on these factors. In that case, a prospective patient population that differs from the historical population in known important factors (e.g., age, weight, duration of disease, or baseline severity) might still be comparable based on a verified ability of the model to accurately adjust for these factors.

The proposed historical dataset is examined with regards to these issues in the following sections.

Historical Patient Comparability to Prospective Study Patients

A central question is whether the patients who have become the focus of the historical dataset are representative of the subjects who will be enrolled in the prospective clinical trial. One, but not the only, aspect of this is the basis of selection for the historical dataset and the prospective study. Two stages of selection are involved. The first is the method by which the patients become known and available to the investigator for screening to assess eligibility. The second are the explicit eligibility

criteria applied during the screening process.   Ascertainment of patients for screening into the historical database required that the Fabry patients come to the clinical centers for evaluation at least once, and preferably for follow-up. Additionally, patients could not be screened unless explicit consent was received, and this was obtained in approximately 58%. The factors that led the patients to present to the clinical centers are not known, as are biases that influenced the fraction in whom consent was obtained. In contrast, in the prospective study Genzyme is actively seeking out Fabry patients. It is unknown in what way the difference between a more passive presentation to the sites over longer timeframes and actively seeking patients will influence the nature of the patients who are ascertained.

The active process may identify patients with lesser severity of disease, or a more mild clinical course. Other patient and disease characteristics may also play a role in determining differences between the populations acquired by passive vs. active processes.

A reasonable assumption is that most Fabry patients will have some period of time when they would be classifiable as Qualified patients. However, only 23% of ascertained patients did become qualifier patients. The largest single reason for not qualifying is the absence of creatinine values within the required range; most non-qualifiers had creatinines too low. The characteristics of patients who returned to the clinical center sufficiently to record values that ultimately reached into the qualifier range compared to those who did not return are unknown. Factors such as disease severity, rate of progression, etc, cannot be discounted.

Another concern for generalizability is the incomplete representation of the available clinical sites and patients. Nearly half (24 of 51) of the sites declined to participate in the historical database collection. The number of patients at these sites is unknown. Again, the differences, if any, between the patients from participating and non-participating sites remains unknown.

Genzyme has attempted to apply similar eligibility criteria to the patients in the historical database and the prospective trial for those patients who were ascertained at the clinical sites. These post-ascertainment eligibility criteria may not be able to equalize differences that occurred in the patient ascertainment, if such occurred. Consequently, there may be differences in the natural history of the two patient sets related to differences in the nature of ascertainment.

An examination of the demographics of the historical database and of patients enrolled into Study 008 noted a difference in age. The mean ages are different between these groups, 37.7 years for the historical database, but 45.5 years for the Study 008 patients at entry (medians 37.3 and 44.8, respectively). The qualified database was examined for a relationship between date of qualification in the database and the age at time of qualification. In any given year of qualification, there is a broad range of ages. However, consistent with the age difference of younger patients in earlier years between the historical database patients and the Study 008 subjects, a suggestion of a modest relationship does exist within the historical database. Median ages are approximately 40 among the last calendar years of patient qualification and in the lower 30's a decade earlier. This difference suggests the possibility that there have been changes in the underlying process leading to patients presenting to clinical centers during the course of the historical database period. Such factors may also be important differences between the historical database patients and the prospective Study 008 also.

The difference in age is not coarsely reflected as a degree of renal dysfunction progression in the Study 008 patients that might be predicted from Genzyme's analysis. The Qualified group in the historical database had a median serum creatinine of 1.3 (mean 1.5) while the Study 008 patients had a median serum creatinine of 1.6 (mean 1.7). The median slope in the historical database Qualified group is 0.086 ln(creatinine)/yr, predicting a difference in creatinine of perhaps 1.1 might have been predicted for the 7.5 year difference in median ages, rather than the 0.3 that is seen.

There are significant differences in nature and frequency of monitoring that occur in normal clinical practice or in routine follow-up in specialized disease-interest clinical centers, compared to the organized, carefully planned evaluations that occur in a prospective clinical trial. The follow-up evaluations in the historical database are much more irregular, and a follow-up visit may be related to a change in status. If this is the case, the data in the database will over-represent patients with significant changes and periods of patient history demonstrating changes (often adverse) in clinical status and under represent the patients who have a largely stable course. This could result in an overestimation of the natural progression rate.

Patients could drop out of the historical database for a variety of reasons that are not random, including disease progression making return to the specialized clinical site (who may well not be their personal physician) too burdensome (when there is no benefit), or lack of interest in returning to the clinical site while experiencing a period of stability (and again perceive no benefit to returning). Which of these effects predominate, if either, or whether there may exist other reasons for drop-out is unknown.

Similarity of Clinical Course of Historical and Prospective Study Patients

Differences in the natural course of patients in the databases may also exist because of differences in medical management over time. At the present time there is no approved treatment for Fabry disease and no standard of practice involving treatments that might be used "off label" for Fabry disease. However, treatments are available for many of the specific complications of Fabry disease. Analgesic medications are widely used, both traditional analgesics and non-traditional treatments often used in neuropathic pain (e.g., certain antiepileptic medications). Substantial numbers of Fabry patients develop and are treated for hypertension. The use of various classes of antihypertensives has changed over the years. Some antihypertensives may lead to better BP control than others, and may provide better nephroprotection.

The historical database includes patients who reached inclusion in the Qualified dataset as early as 1972, and some who were included based on evaluations made as late as 2002. The majority of patients included had the start of the qualified period within the past 10 years (87 of 103 in 1992 or later).

The dataset was examined for an influence of the year of qualification on the estimated slope. No consistent relationship was found between these two values.

Ultimately, as for the previous concerns, the influence of medical management changes over the years on the progression of renal disease in Fabry patients cannot be formally tested.

Robustness of the Historical Data

Inspection of the Patient Data

As a coarse overview of this dataset, the data from individual patients can be examined and a general impression of the quality of the data and degree of linearity of ln(creatinine) vs. time can be gained. In assessing this it should be born in mind that the progression criterion of a 50% rise in creatinine translates to a rise of 0.405 in ln(creatinine). Deviations from linearity that are of 0.1-0.2 ln(creatinine) units constitute a substantial error from the model compared to the success/failure criterion. Deviations that lead to a slope shift of just 0.05/yr or less can be the difference between a patient projected to show a moderate, non-endpoint-achieving rate of increasing dysfunction to one projected to exceed the progression criterion within 3 years.

Selected example patients are shown in the following figure. Section A illustrates 3 sample patients that appear to follow the linear assumption reasonably well. Although the data are of shorter duration than the 3 year study period, they are adequate to discern that these patients would be readily regarded as renal progressions.

Section B illustrates a patient that appear to not conform to the linear rise assumption. This patient appears to have a biphasic course consisting of a slower progressing period for several years followed by a more rapid rise. A rise in creatinine by approximately 1.6 natural log units (indicating an increase of 500%, a factor of 5 in creatinine) occurs, much of which is just after 3 years past qualification. A linear fit to these data will be strongly influenced by this late, large increase. Section C illustrates another patient with an apparent non-linear course. Although the total rise in ln(creatinine) is less, it occurs after a much longer period of stability (note different time scale).

Section D again shows a patient who had a non-linear rise, with the increase seen only at 4.5 years past qualification. However, for this patient there is a substantial gap in the data record (almost 2.5 years), and the data do not permit identification of the actual time of conversion to rapid increase. Where the stable period ends and the rise begins is unknown, and might extend beyond 3 years. However, a linear fit to the data will again be influenced by this large, post-2-year-gap rise in creatinine. Section E also illustrates a patient with a long gap in creatinine values (6 years duration) after which creatinine has begun to rise. Again, where the rise begins, in particular before or after year 3, is unknown.

The impact of some of these types of potentially difficult data will be examined in sensitivity analyses in the following sections.

Section F shows an example of a patient where the majority of the data illustrate a pattern that is potentially confounded by a single value which appears to be physiologically unlike the adjacent values. However, strict line fitting which makes some attempt to conform the line to the data will shift the estimated slope to a greater value, and give an impression of the patient's course that is not conveyed by examination of the data.
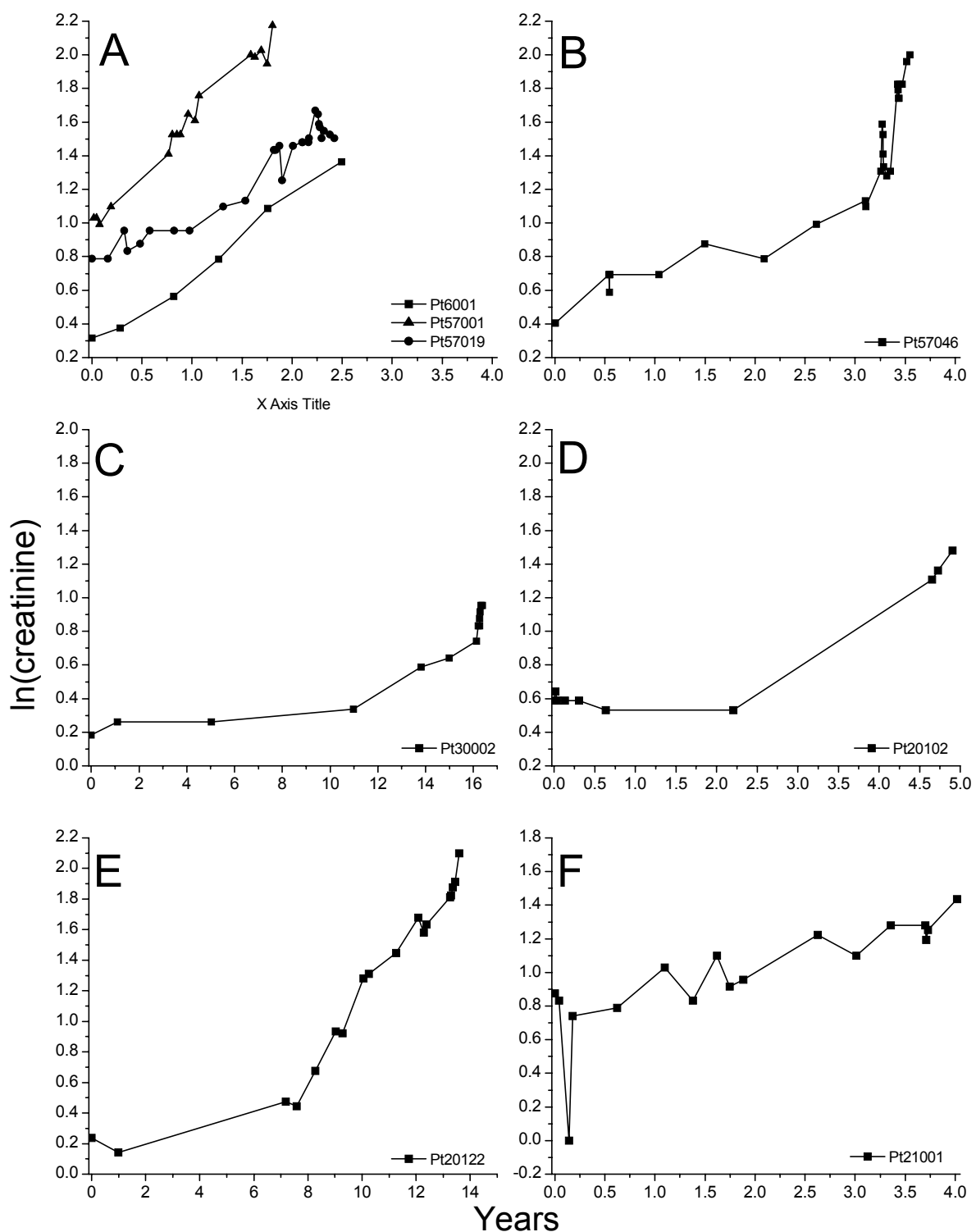
1)   Note differences in time scales, particularly panels C and E.
2)   In panel A, data of Patient 6001 shifted to lower ln(creatinine) by 0.2 at all points to provide data separation
      assisting readability .

These represent only 8 patients of the total 63 with 3 or more data values. There are also patients with only 2-3 points over a 1-2 month period who appear to show increases of creatinine by 0.1 or 0.2 , which translates to changes in ln(creatinine) of $\geq$ 0.05 /1-2 months, or estimated slopes of $\geq$ 0.3 /yr. The strength of extrapolation of such data to 3 years is uncertain.

While this approach is useful to gain an overview of the nature of the data, examination of individual patients is a potentially biased and highly subjective method to formulating judgments. More comprehensive approaches can be less subject to selection bias.

The median period of follow-up data is only 1.4 years; half the patients have data describing a timeperiod that is less than half as long as the observation period of the proposed trial. There is one month or less of follow-up on 41 patients (40%), and 78 patients (76%) have 4 years or less of follow-up data. Overall, there is a tendency for the patients with the most datavalues to have a longer duration of qualified-period follow-up.

Comprehensive assessments of robustness

Genzyme proposes to use the linearly modeled historical database to project the renal progression-event rate over 3 years, with a progression-event defined as a 50% increase in creatinine. A robust data set and analysis should not be influenced unduly by patient experience that is well beyond that which would be observed in the prospective study. Neither should the analysis be overly influenced by large gaps in observation (that would not occur in the prospective study) with implied imputation of patient experience during that gap. Several explorations of the data were performed to assess the sensitivity of the analyses to more extreme values of creatinine or to potentially important gaps in the database due to unplanned irregular follow-up. Two analyses were devised to test the robustness of the dataset to provide stable estimates of the progression rate, one presented here and one in the next section.

The first analysis addresses the fact that the model will be applied to a prospective clinical trial wherein the clinical course of creatinine levels only up to a rise of 50% will be used to categorize a patient as a success or failure. However, all available data are used from the historical dataset, including data of much higher rises in creatinine. These much larger rises may unduly influence the estimate of the rate of rise during the earlier portion of the patient's course. Therefore, this analysis drops out all data in each patient's series of values that show or follow a finding of a creatinine rise of more than twice the "baseline" value. While twice the baseline value is substantially higher than is relevant for the prospective clinical trial, this was thought to be an adequately judicious criterion that deleted only data clearly not informative as to the time course of more modest rises. An important note is that in a dataset planned for this modeling method, there would be substantial numbers of creatinine values between the baseline and twice the baseline from which to estimate a slope. Consequently, the expectation would be no change in the slope estimate when this reduced dataset is used for the analysis.

This analysis (drop values occurring after a creatinine rise to more than twice baseline) eliminated 75 values among the 103 patients, leaving 103 patients with 508 values from the original 583 values. Thus, 87% of the datavalues were retained. Modeling based on this reduced dataset notably changes the estimate (illustrated in the following figure showing the cumulative slope distribution).

Estimated Slope Distribution with Exclusion of Data Beyond Doubling of Creatinine
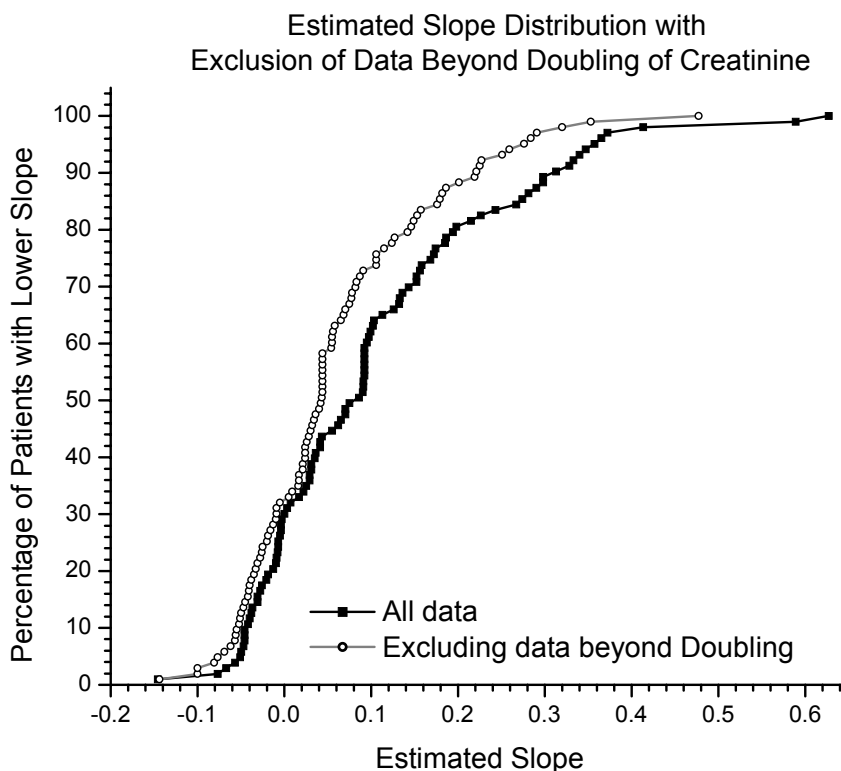
This analysis shows little change in the curve for patients with estimated slopes of only slow changes. However, in the region of critical interest (percentage with slope > 0.135), there are notable differences in the curves. This analysis suggests that only 22 of 103 patients have an estimated slope of greater than 0.135, projecting a 21% progression rate, as compared to the 32% progression rate from the data set with all datavalues. This 1/3 change in the estimated progression rate resulted from removal of only the 13% of data values that are well beyond the creatinine rise of interest. This analysis illustrates that the proposed analysis method is highly sensitive to the more extreme values of creatinine in a patient's data. For a substantial number of the patients estimated to be "progressing patients" the slopes were heavily influenced by the presence of the creatinine values of more than twice baseline.

Genzyme performed an alternative sensitivity analysis in which instead of excluding data values, all data showing a more than doubling of creatinine are replaced with values indicating a doubling. This dataset has 583 values as does the original, but with 75 altered values. Analysis of this dataset indicates a 25% projected progression rate (26/103 patients with slopes > 0.135). This intermediate value is not surprising, since it restores to the dataset creatinine values at the highest end of observed values.

This reduced dataset sensitivity analysis does not demonstrate that the "correct" estimate of progression rate is the lower one produced by the sensitivity analysis. While it may be a better dataset for this analysis by allowing the earlier time periods with more relevance to the proposed clinical study to have greater influence, it may still not be a fully accurate projection. However, the true rate of progression might be as low as, or even lower than the 21% seen in this sensitivity analysis. Importantly, the analysis does suggest that the available historical dataset is not robust with regards to the proposed method of analysis and interpretation.

Linearity of Logarithm of Creatinine Change Over Time

Lowess Curve Analysis

Genzyme concluded that linear fits to the data are appropriate, based in part on a lowess curve analysis. The lowess curve is a moving average estimate of the mean, where observations closest to the center of the moving window are weighted more heavily than observations lying further from the center of the moving window. Genzyme excluded data beyond 5 years because few patients have data after 5 years, and these patients could unduly influence the lowess estimate of means at early time points. Also, renal events in the prospective study are defined based on 3 years of observation for each patient, and measurements beyond 5 years are less relevant. The lowess method employs a window-width parameter which can affect the apparent smoothness of the curve. In choosing the window-width, there is a trade-off between bias and variance. The wider the window, the smaller the variance of the estimated curve at any fixed time since more observations are being averaged. However, estimates which are smoother than the functions they are approximating are biased.

CBER Biostatisticans have examined the effect of the window-width. The following figure shows the different estimated curves using the lowess method when different window-widths (bandwidths) are used. The estimated curves in the two upper and the lower left were fit using narrower bandwidths than that used by Genzyme. The estimated curve in the lower right is the same as that supplied by Genzyme. This figure illustrates that the relationship of log serum-creatinine over time depends on what bandwidth is selected in the lowess method. The estimated curve is not necessarily linear, but dependent upon an arbitrary selection of the window-width. Therefore, this analysis does not support the conclusion that the assumption regarding linearity of log serum-creatinine over time is justified.
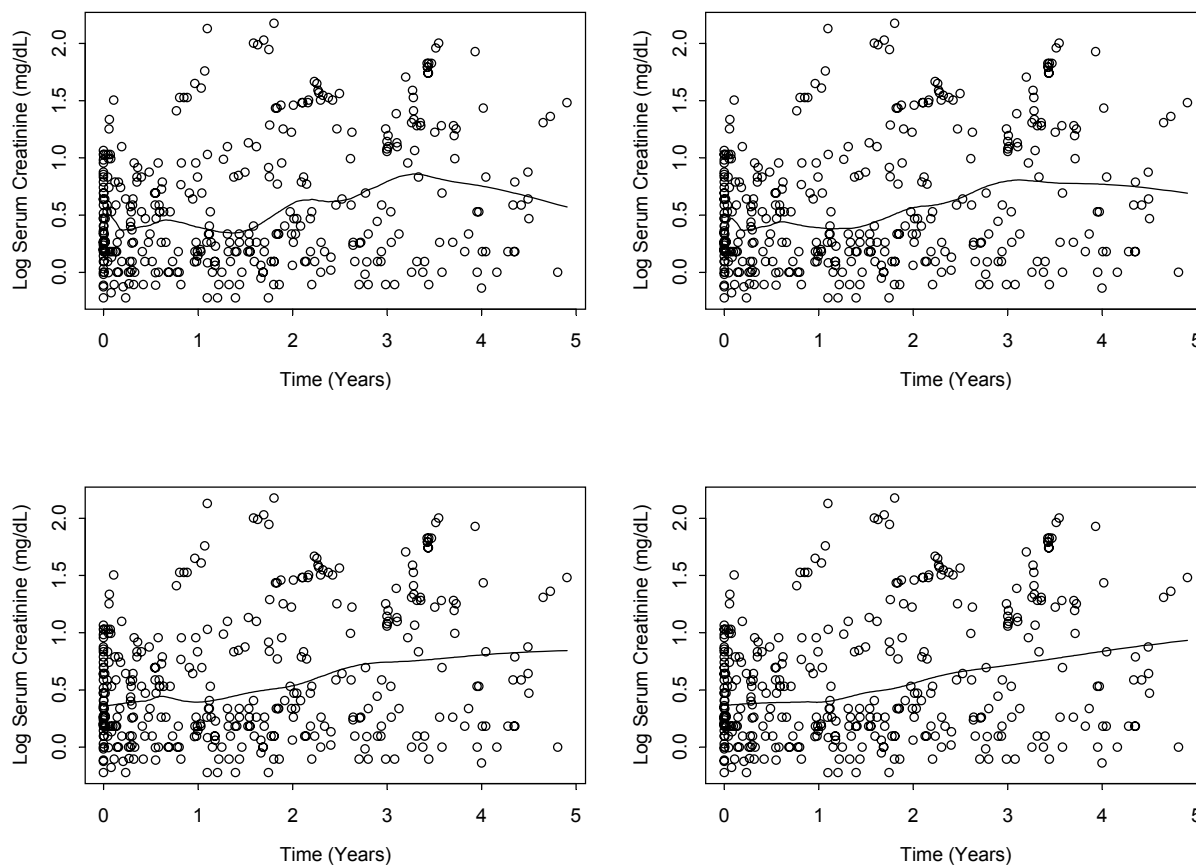
A lowess curve analysis may not be the best method to assess linearity. This approach emphasizes the data direction of the entire dataset over time. Genzyme posits that each patient may have a unique linear rise, rather than all patients having the same rate of rise. There is a broad range of different frequencies and total durations of data collection in the dataset. Thus, any narrow window-width will capture data from a varying composition of patients as the window moves. The expectation of non-identical slopes for patients suggests that the lowess curve approach may be unable to address the question of linearity within each patient.

Residual Differences with Linear and Quadratic Models

A different approach to assessing the linear fit is to consider the residual differences between the model and the data. In a perfect fit to data, there will be no difference between the modeled estimate and the actual data at each timepoint for which a data value occurs. When the data are noisy, or the model does not fit well to the data, there will be "residual" differences between the data and model values. Comparisons of the summed squares of the residuals is one way to assess the relative quality of fit between two models.

To further assess how well a linear model fits the data compared to a higher order fit, standard fits of constant, linear, and quadratic models were made, and the sum of squares of residuals compared. For this comparison of models, the empirical Bayes method is not appropriate, since that method induces the fit-estimate to deviate from the best fit for the individual patient data in order to conform better to the entire patient population. Consequently, it will not produce the lowest possible residuals within each patient's data. Because quadratic equations can perfectly fit 3-value sequences and could be considered a biased assessment, only patients with at least 4 data values were used for this analysis (n=54 of the 103 total). Sum of squares of residuals (SSR) for differing models were evaluated by ratio of the SSR of two models (e.g., SSR-linear/SSR-constant).

Ratios of Sums of Squared Residuals				
Fitted models	mean ratio	Lower Quartile	Median	Upper Quartile
Linear vs Constant	0.512	0.221	0.438	0.786
Quadratic vs Constant	0.339	0.122	0.255	0.543
Quadratic vs Linear	0.633	0.391	0.702	0.949
Difference in SmSqRsd between Linear - Quadratic divided by SmSqRsd of Constant	0.172	0.013	0.108	0.284

n=54 all comparisons; only patients with 4 or more creatinine values

This analysis illustrates that, as expected from the fact that renal dysfunction does progress, the linear fit is a substantial improvement over the model of a constant value (i.e., no change over time), reducing the summed squared residuals to a median of 44% of the no-change model. However, the quadratic model further reduces the summed residuals as well, to a median 70% (average 63%) of the linear change model. While this is not as large a reduction overall as moving from constant to linear models, this indicates that a model that allows for curvature of the change over time is a better approximation to the data than the linear fit. Not all patient's data are a markedly better fit with the curved fit, but the data of many of the patients demonstrate curvature and a notable reduction in the residuals with curved fits.

Genzyme has acknowledged that the quadratic term is statistically significant in these fits, and that the quadratic model is a better fit of the data. An important note on these curved models is that these analyses do not demonstrate that the quadratic equation is the ideal fit for the data, nor that it is appropriate to use. Rather, the improvement with the quadratic fit indicates only that there is an important aspect of curvature to the within-patient data over time.

Genzyme applied the quadratic model (applied to the 104 patient Final dataset) and reported an estimate of a 29% progression rate, compared to the 32% rate of the linear model. Thus, the quadratic model method suggests a small, but non-negligible difference in progression rate.

Sensitivity to the Data Transformation

Recognizing that creatinine values do not rise linearly over time as renal impairment progresses, Genzyme selected the log transformation to attempt to linearize the data. This is an arbitrary

selection, without any substantive biologic justification (see above, Assessment of Stability of the Projected Rate section). Other data value transformations can also be examined. Inverse creatinine (1/creatinine) is a transform that has often been used in examining rates of renal insufficiency progression. Upon FDA request, Genzyme performed linear modeling with the inverse creatinine transform to the data (the 104 patient Final dataset). This provided an estimate of 23% progression rate, compared to the 32% for the ln(creatinine) transform method.   While this analysis does not assess the quality of the linearity of the transformed data, it illustrates that the linear modeling method is not robust to different, arbitrarily selected transformation methods.

Slopes Across Patient Subsets by Initial Creatinine

In another sensitivity analysis of linearity, the starting value for each patient (the creatinine at the date of qualification, analogous to the baseline value in a prospective study). Since a line has the same slope irrespective of which point along the line is treated as the starting point, the slope estimate for each patient should be unrelated to which point in time along their history of progressive renal dysfunction they begin to have records at the clinical site that designates them as qualified.

The data of ln(creatinine) at qualification ("baseline" creatinine) was divided into cohorts of width 0.1 unit and each cohort analyzed. This analysis does not depend on "outcome" data anymore than an analysis of baseline creatinine in a prospective study would. A relationship between the baseline creatinine and the estimated slope was observed, as illustrated in the following figure.

Relationship of Baseline Creatinine to Slope
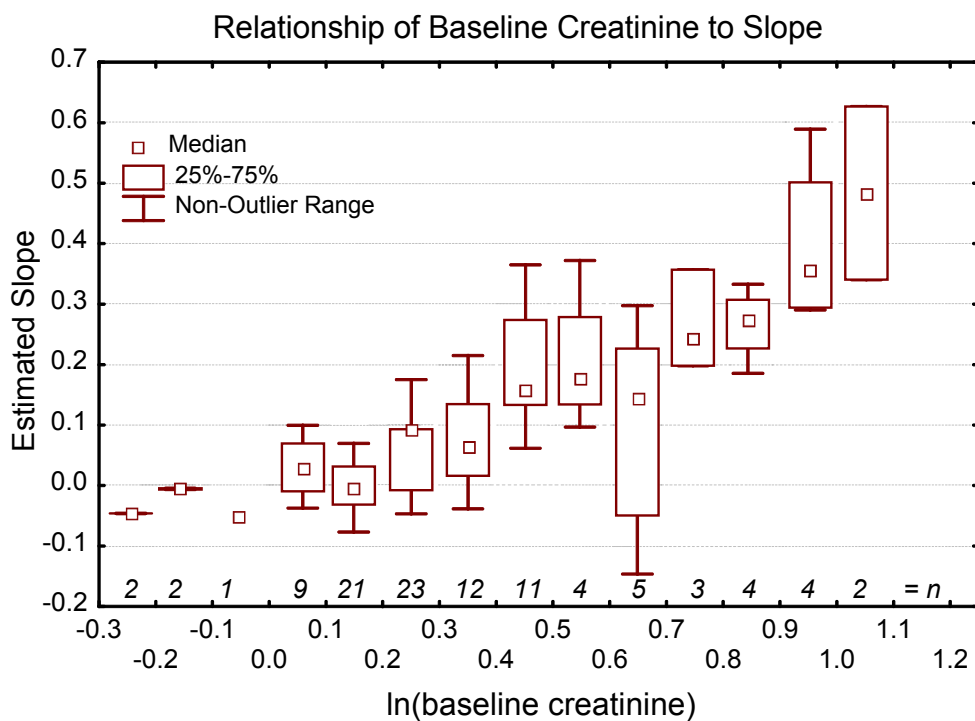
The results illustrated in this figure strongly suggest that the assumption of a linear rise in ln(creatinine) over time is invalid. However, if the renal dysfunction progresses in an accelerating rise of ln(creatinine) then the rate of rise at baseline will vary with how advanced the patient's renal dysfunction is at the date of entry to the database. Thus, this analysis suggests that there is an accelerating rise in ln(creatinine) over time, rather than a constant linear rise. While the data are sparse in the groups near the ends of the creatinine range, they nonetheless support the pattern of the central groups. These data are too sparse to permit precisely defining the relationship between creatinine at time of qualification (and therefore, at study enrollment) and slope, but strongly suggest that it is not immaterial.

Reduced Dataset Due to Creatinine Rise During Extended Time Gap

In addition to the sensitivity analysis of dataset robustness described above in which data were dropped if the creatinine rise was substantially in excess of the relevant amount, a related second sensitivity analysis was also conducted,. This additional analysis examines the sensitivity of the dataset analysis to long gaps between creatinine values when dramatic rises occur.

In this second sensitivity dataset considers the problem that substantial gaps in the data where dramatic rises of creatinine occur do not allow one to differentiate between linear rise across the entire gap period (which is assumed by the model) and stable creatinine levels for a portion of the gap-period, followed by a period of sharper rise (biphasic, or accelerating curved rise). Panels D and E (page 19) show patients with important gaps in the creatinine record. Therefore, a sensitivity dataset was formulated where data were deleted from each patient's series of values if they show or follow a creatinine rise of more than 50% from the prior value and that rise occurred during a time-gap of 1 year or more.

This dataset is biased because it removes data that show marked increases in ln(creatinine). However, it does not eliminate all rapid-rising values, only those where there are important weaknesses in the dataset; i.e., where the point in time of increase is highly uncertain. In an ideal dataset, there would be frequent monitoring, and no long gaps in time. Patients where the gap in time is less then 1 year will have no data eliminated. Where gaps are 1 year or more, the uncertainty in the time of rise is 1/3 of the duration of the proposed prospective study or more. Gaps of several years (e.g., Panel E, page 19) are longer than the proposed study. Furthermore, there is no need to eliminate long gaps if the creatinine rise has not been dramatic. If the creatinine is unchanged or only modestly higher at the end of the gap, an assumption of a linear course during the gap is unlikely to be incorrect.

This criterion (drop values after a sharp creatinine rise during a gap of 1 year or greater) eliminated only 31 values among the 103 subjects, leaving 552 datavalues (95% of the data). No patients are eliminated in this sensitivity dataset. The results are shown in the following figure.

**Distribution of Estimated Slopes with
Exclusion of Data Beyond Gaps of 1yr with 50% Rise**
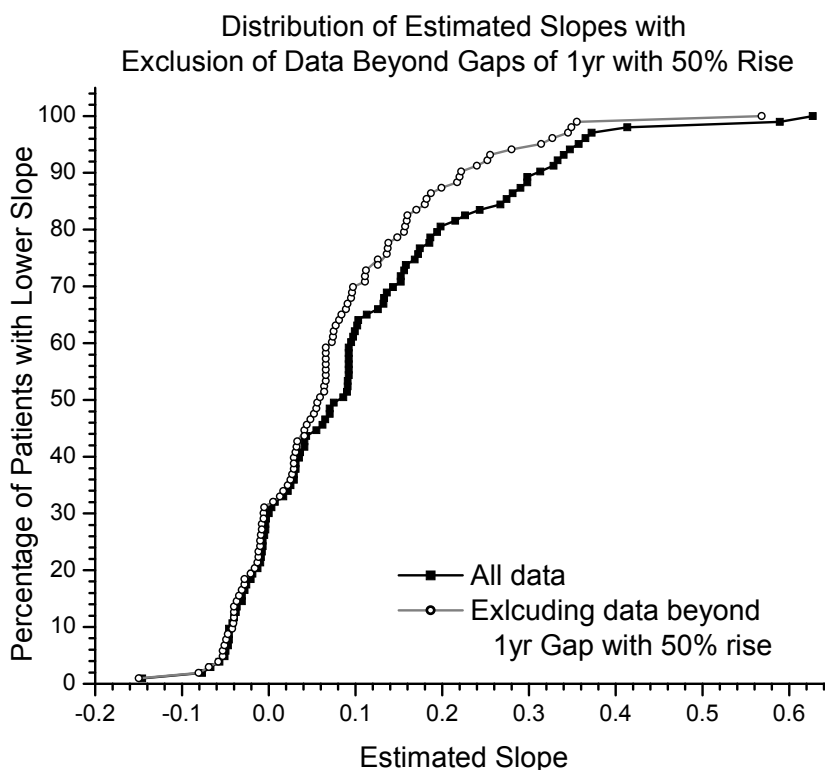
This analysis shows little change from the complete dataset for patients with low slopes of change. However, again in the region of critical interest (slope = 0.135) the curves diverge. This analysis suggests that only 26 of 103 patients have an estimated slope of greater than 0.135, projecting a 25% progression rate, as compared to the 32% projection of the complete dataset. Thus, elimination of the 5% of the data which represents uncertainty as to time of change within a patient led to a notable decline in the estimated progression rate (a 22% relative decrease). Since a truly linear course of progression would not be so sensitive to removal of a small amount of data, this analysis also brings into question the appropriateness of the linearity assumption. Equally importantly, this analysis also demonstrates that this dataset is not a robust dataset for this type of analysis.

A further analysis was run applying both reduced dataset sensitivity alterations together. This resulted in a dataset of 103 patients with 489 values compared to the original 583 (84% of the data retained). As can be expected from the prior two sensitivity analyses, this analysis also resulted in a markedly reduced estimate for the rate of patients progressing by 50% within 3 years (19%). None of these sensitivity analyses can be regarded the definitive accurate estimates of the progression rate. Rather, they do raise doubt of the dataset's robustness.

Adequacy of the Random Effects Model with Empirical Bayes Estimation As Proposed
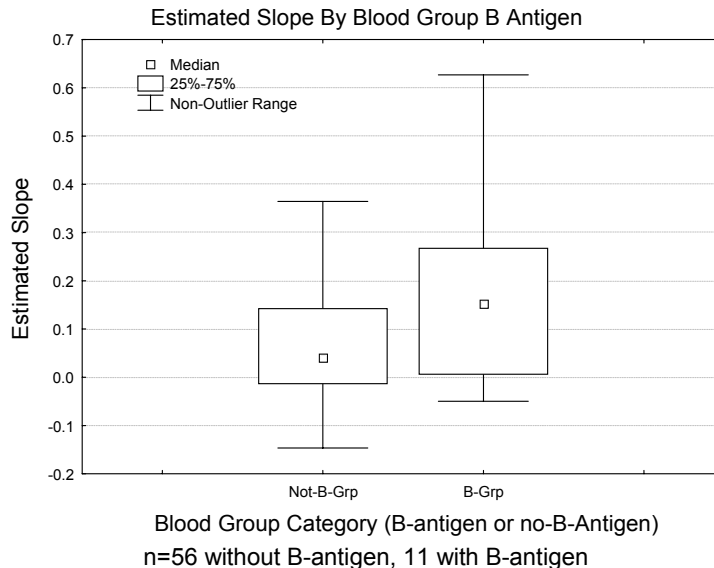
Genzyme proposed that no factors present at the time of qualification (analogous to the demographic and baseline characteristics at the time of study enrollment) need to be considered in

order to model the data well. Several factors of potential relationship to the rate of renal dysfunction progression were examined to assess this assertion.

Age was previously noted to be different between the historical database and the patients currently enrolling into Study AGAL-008. There was also a modest relationship between the year of qualification and the age at qualification.   However, when the estimated slope was examined by age in 5-year wide cohorts, there was no notable trend in the mean slope across the age cohorts.

There are few women in the qualifier historical database (as in the AGAL-008 study) which precludes conclusions about the effect of gender.

Fabry patients who have blood group B antigen have accumulations of blood group B glycosphingolipids in addition to two other glycosphingolipids that accumulate in all Fabry patients. Consequently, blood group B Fabry patients may have more severe or rapidly progressive disease. There were many missing values in this database regarding blood group; only 67 of the 103 qualified patients had this information. When the relationship between blood group and slope was examined (see figure below) in the database the B-antigen positive patients had a mean slope of 0.172 $\pm$ 0.060 (n=11, median 0.152) while B-negative patients had a mean slope of 0.085 $\pm$ 0.140 (n=56, median 0.042). This suggests that B-antigen may be an important predictor of the rate of progression, and with a strength of influence that may be important relative to size of the critical slope of 0.135. However, these data are too limited to reach a definitive conclusion, or to precisely estimate an adjustment factor for blood group.

Estimated Slope By Blood Group B Antigen

Blood Group Category (B-antigen or no-B-Antigen)
n=56 without B-antigen, 11 with B-antigen

As discussed previously, the Empirical Bayes estimation method allows slope estimation for patients with even a single creatinine value. In addition, slopes for patients with relatively fewer datavalues will be influenced by slopes from patients with greater numbers of datavalues. In order to assess the impact, the distribution of slopes was also examined in a method where the linear modeling of increase in ln(creatinine) with time was retained, but the line fitting done without the

empirical Bayes estimation. This model was calculated on the 85 patients with 2 or more datavalues. The resulting cumulative distribution of slopes is shown in the following figure.

Distribution of Estimated Slope with Different Modeling Methods
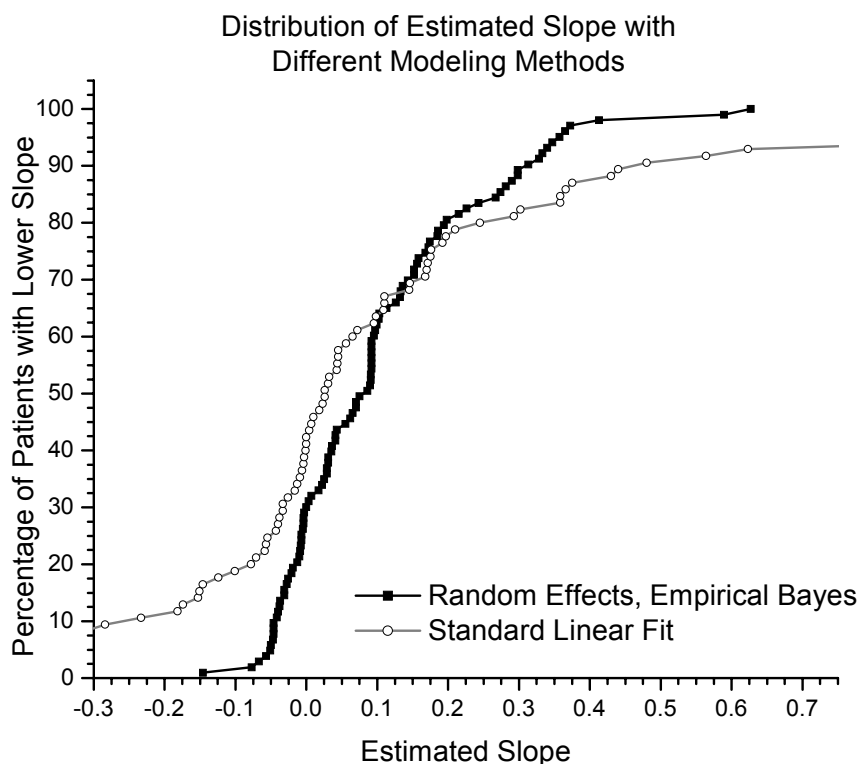
This figure illustrates clearly the impact of the random effects model with empirical Bayes estimation.   The distribution of slopes is notably sharper (less varied) with empirical Bayes estimation than in the standard linear model. The standard linear model provides an estimate of 33% of patients (28/85) with slopes greater than 0.135, compared to 32% for the REEB method. The two methods are fairly close in their distribution in the slope range of approximately 0.1 to 0.2 (approximately 60th to 80th percentiles).

Note that the similarity of REEB method and standard modeling does not pertain if the quadratic modeling method is used. As shown earlier, the data appear to have an important element of curvature, and quadratic modeling fits the data better than linear. If quadratic modeling is applied to the dataset for patients with 3 or more datavalues (so that quadratic fits are possible; n=63) the standard (each patient independently) method estimates that 41% (26 of 63) will have progression compared to 31% with REEB quadratic modeling on the set of 64 patients with 3 or more datavalues (in the Final dataset).

Summary

Genzyme conducted an extensive international effort in the collection of natural history data of Fabry patients, and analyzed the creatinine data. Genzyme proposes to use these data as a control

for comparison to a prospective single arm Fabry disease treatment trial with galactosidase. This dataset is the largest, and most comprehensive database of Fabry patients known to either FDA or Genzyme.

Genzyme proposes to use this database to model the course that untreated Fabry patients would have in a prospective study if there was such a group. Genzyme models the data with a linear model, and derives an estimate of the slope for each patient using a random effects with empirical Bayes estimation. The set of slopes was then examined for the fraction of patients who are expected to show an increase of creatinine by 50% from baseline within 3 years (defining a progressor patient).

However, the set of "qualified" patients is derived from a small minority of all Fabry patients identified at the participating sites (only 58% of identified patients did participate, and only 23% of participating patients have "qualified" periods of data), and only about half of the invited sites participated. Most patients have few datapoints (40 with only 1 or 2, and 58 with 4 or fewer), and very few patients have a duration of valid data in the dataset long enough to permit renal progression status to be determined over the relevant 3-year time course for the proposed clinical study.

Consequently, this modeling approach is dependent upon implied extrapolation and interpolation of the data to yield a categorization of the progression status for each of the qualified patients.   From this progression status the group progression proportion is determined.

Genzyme calculated a point estimate of 32% for the portion of patients expected to progress by a rise of 50% in creatinine within 3 years of study entry. There are substantial flaws in the method used to provide a range of progression rates proposed as the confidence interval. The manner of constructing a confidence interval for the percentage progressing remains a difficult issue without a valid resolution at this time. Importantly, any confidence interval calculated from the dataset reflects some types of uncertainty but does not account for other important types of uncertainty or bias (e.g., uncertainties in the representativeness of the population, laboratory errors, changes in the population presenting to the clinical centers).

Certain issues regarding the historical population remain poorly assessed. While there are no major changes in the management of these patients known to have made the clinical course of patients of a decade ago unlike those of the present time, there is nonetheless no way to assure the historical database patients are comparable to patients enrolled into a prospective study.

The influence of selection bias cannot be determined. It is unknown if the small fraction of patients who are included as "qualified" patients is in some way biased from the larger population of all Fabry patients, and how such biases to participation may or may not operate in the recruitment of patients for a prospective treatment trial.

The assumption of linearity was examined in a number of ways. Examination of the individual patient data indicates for some patients linear rise in creatinine is not a good model. More comprehensive analyses suggest that the logarithm of creatinine is not simply a linearly rising function over time. Instead there appears to be a significant component of curvature to the rise.

The modeling method proposed by Genzyme may not provide reliable estimates of the population progression rate. Additionally, a modeling process that ignores the baseline creatinine level to calculate estimates of predicted group rate of progression will not be well suited for comparison to another group that has a different distribution of baseline creatinine values.

The blood group of the patient may influence the rate of renal progression. However, the available data are too limited to develop precise estimates of the influence. Thus, a model that incorporates adjustment for this or other factors cannot be developed.

The robustness of the critical characterization of patient progression status is a centrally important feature of a good natural history database. The dataset was found to not be robust to inclusion/exclusion of either extreme values of creatinine well beyond those that could be observed in the proposed clinical trial, or to inclusion/exclusion of gaps in data collection at times when marked increases in creatinine occur. Removal of a minority of the datavalues markedly altered the projected progression rate.

Genzyme's process of interim analyses also illustrate the unstable nature of the dataset. The estimate based on the first portion of qualifier patients is notably different than that based on the second portion of patients.

The estimated progression rates from the different methods of generating estimates are summarized in the following table:

Summary Listing of Progression Rate Analyses					
Fit Method	Fit Order	Explanatory Note	Number of Patients	Number of datavalues	Estimate of Progression Rate
Random Effects, Empirical Bayes	Linear	Genzyme proposed method	103	583	32%
Empirical Observation	none	Patients with outcome value within 2.5 - 3.5 years	17	n/a	41%
Random Effects, Empirical Bayes	Linear	Early collected patients	43		40%
Random Effects, Empirical Bayes	Linear	Later collected patients	60		27%
Random Effects, Empirical Bayes	Linear	All patients, data censoring after creatinine doubles	103	508	21%
Random Effects, Empirical Bayes	Linear	All patients, data censoring for long time gaps with marked creatinine rise	103	552	25%
Random Effects, Empirical Bayes	Quadratic	All patients	104		29%
Random Effects, Empirical Bayes	Linear	1/creatinine transform	104		23%
Random Effects, Empirical Bayes	Linear	subset by baseline creatinine	104		Strongly related
Each Patient Independently	Linear	Patients 2 or more values	85		33%
Each Patient Independently	Quadratic	Patients with 3 or more values	63		41%
Random Effects, Empirical Bayes	Quadratic	Patients with 3 or more values	64		31%

The differences between these estimates are substantial when viewed in the context of how they would be used. Even a 5% difference in rate for comparison to an observed rate of progression in a single arm treatment protocol has the potential to be very important to both sample size calculation and interpretation of observed results.

Consequently, no well founded value for the projected progression rate can be derived from this dataset. Selection of an expected control rate that is too high will lead to Type I errors and selection of a control rate that is lower than actual will dispose to Type II error.

The non-robust nature of the dataset for this proposed use, the inability to adequately determine and adjust for important baseline characteristics, and the absence of an understanding of the range of uncertainty in the estimated progression rate are serious flaws in the proposed approach to a verification study. This dataset does not permit a precise estimate of the expected progression rate.

Recommendation

The modeling method proposed by Genzyme is unable to yield a robust value for estimation of the progression rate. FDA concluded that their approach is unsuitable for use as a historical control in a

clinical trial. Any other method based pm this dataset will be subject to many of the same difficulties, and highly dependent upon extrapolation and interpolation to form projections for comparison between trials. No modeling method has been validated to provide the necessary extrapolation and interpolation of creatinine values. Therefore, FDA cannot endorse any of these methods as being adequately rigorous.

A better founded estimate of renal progression rate can be derived from empirical observation on a less sparsely informative dataset. This current dataset has only 17 patients who have adequate data to form an empirical assessment. The reason for the sparseness is that data were collected only from the specialized centers that saw larger numbers of cases.   These site investigators were often not the primary care physicians for these patients, and the patients may have been seen infrequently. However, it is likely that the patients were seen for more regular follow-up by their personal physician.   Thus more data on serum creatinine values for these patients should be available from the medical records of their personal physicians. If the sponsor wishes to pursue the historical comparison method further, it is FDA's view that Genzyme should seek to collect all serum creatinine values from all sources.   If the database were made more complete, then there is a possibility there would be adequate data to permit the straightforward empirical observation method to provide a more robust estimate for progression rate.

Addendum: New Genzyme Proposal for Historical Comparison

Genzyme recently developed an outline for a new method for the historical comparison approach. This method has not been submitted to FDA with adequate time for review prior to the deadline for finalization of this document.   Therefore, FDA is unable to provide sustentative comment at this time.

A brief overview of the concept is that a technique termed Propensity scoring will be used to select patients and "initial creatinine" timepoint from the historical dataset for use in comparison with the agalsidase treated patients. Propensity scoring will assist in best matching each historical control patient to a patient in the Study 008 based upon certain selected covariates (it does not, however, seek to ensure that each Study 008 patient has a good match provided).

This new proposal recognizes that a potentially valuable second comparison group may be feasible from the existing placebo patients in Study 008. There are 73 patients enrolled into this study, with a randomization ratio of 2:1 (agalsidase:placebo). Enrollment began in January of 2001, and was complete in August of 2002.   Thus, in January 2003 the earliest patient will have 2 years on study, and the median-enrollment (Nov. 2001) patient will have 14 months on study. It is notable that Genzyme's design for Study 008 proposed that the study (in the fully randomized, placebo controlled form) would thus be complete in January 2004.

The proposal recognizes that there will be missing data in the historical patient dataset, and in the time-censored (due to early termination of the placebo control) placebo patient data. The pattern of missing data is expected to be complementary between the two groups. In this new method imputation of missing data in each of these comparison groups will be performed by drawing upon

the data in the other group. The exact method for this has not been stated. The exact nature of the imputation method is obviously important; it may still be subject to many of the weaknesses of the implied interpolation and extrapolation modeling methods discussed above.

After imputation (proposed to be performed as multiple imputation in practice) the "filled-in" control data set will be used to generate a prediction of outcome for the agalsidase treated patients if they had been untreated. Again, the prediction technique is left unstated at present in the proposal. The actual observed outcome in the agalsidase patients is then compared to the prediction.

Thus, this method attempts to address some of the weaknesses of the prior proposal. Not all sources of uncertainty described above are addressed. Futhermore, the proposal is incomplete, with key elements of the method remaining undefined at present.